

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/116045>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Low-rank matrix recovery with Ky Fan 2- k -norm^{*}

Xuan Vinh Doan^{1,2} and Stephen Vavasis³

¹ Operations Group, Warwick Business School, University of Warwick, Coventry,
CV4 7AL, United Kingdom

`Xuan.Doan@wbs.ac.uk`

² The Alan Turing Institute, British Library, 96 Euston Road, London NW1 2DB,
United Kingdom

³ Combinatorics and Optimization, University of Waterloo, 200 University Avenue
West, Waterloo, ON N2L 3G1, Canada
`vavasis@uwaterloo.ca`

Abstract. We propose Ky Fan 2- k -norm-based models for the non-convex low-rank matrix recovery problem. A general difference of convex algorithm (DCA) is developed to solve these models. Numerical results show that the proposed models achieve high recoverability rates.

Keywords: Rank minimization · Ky Fan 2- k -norm · Matrix recovery.

1 Introduction

Matrix recovery problem concerns the construction of a matrix from incomplete information of its entries. This problem has a wide range of applications such as recommendation systems with incomplete information of users' ratings or sensor localization problem with partially observed distance matrices (see, e.g., [3]). In these applications, the matrix is usually known to be (approximately) low-rank. Finding these low-rank matrices are theoretically difficult due to their non-convex properties. Computationally, it is important to study the tractability of these problems given the large scale of datasets considered in practical applications. Recht et al. [11] studied the low-rank matrix recovery problem using a convex relaxation approach which is tractable. More precisely, in order to recover a low-rank matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ which satisfy $\mathcal{A}(\mathbf{X}) = \mathbf{b}$, where the linear map $\mathcal{A} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^p$ and $\mathbf{b} \in \mathbb{R}^p$, $\mathbf{b} \neq \mathbf{0}$, are given, the following convex optimization problem is proposed:

$$\begin{aligned} \min_{\mathbf{X}} \quad & \|\mathbf{X}\|_* \\ \text{s.t.} \quad & \mathcal{A}(\mathbf{X}) = \mathbf{b}, \end{aligned} \tag{1}$$

where $\|\mathbf{X}\|_* = \sum_i \sigma_i(\mathbf{X})$ is the nuclear norm, the sum of all singular values of \mathbf{X} . Recht et al. [11] showed the recoverability of this convex approach using some

^{*} This work is partially supported by the Alan Turing Fellowship of the first author.

restricted isometry conditions of the linear map \mathcal{A} . In general, these restricted isometry conditions are not satisfied and the proposed convex relaxation can fail to recover the matrix \mathbf{X} .

Low-rank matrices appear to be appropriate representations of data in other applications such as biclustering of gene expression data. Doan and Vavasis [5] proposed a convex approach to recover low-rank clusters using dual Ky Fan 2- k -norm instead of the nuclear norm. Ky Fan 2- k -norm is defined as

$$\|\mathbf{A}\|_{k,2} = \left(\sum_{i=1}^k \sigma_i^2(\mathbf{A}) \right)^{\frac{1}{2}}, \quad (2)$$

where $\sigma_1 \geq \dots \sigma_k \geq 0$ are the first k largest singular values of \mathbf{A} , $k \leq k_0 = \text{rank}(\mathbf{A})$. The dual norm of the Ky Fan 2- k -norm is denoted by $\|\cdot\|_{k,2}^*$,

$$\|\mathbf{A}\|_{k,2}^* = \max_{\mathbf{X}} \langle \mathbf{A}, \mathbf{X} \rangle \quad \text{s.t. } \|\mathbf{X}\|_{k,2} \leq 1. \quad (3)$$

These unitarily invariant norms (see, e.g., Bhatia [2]) and their gauge functions have been used in sparse prediction problems [1], low-rank regression analysis [6] and multi-task learning regularization [7]. When $k = 1$, the Ky Fan 2- k -norm is the spectral norm, $\|\mathbf{A}\| = \sigma_1(\mathbf{A})$, the largest singular value of \mathbf{A} , whose dual norm is the nuclear norm. Similar to the nuclear norm, the dual Ky Fan 2- k -norm with $k > 1$ can be used to compute the k -approximation of a matrix \mathbf{A} (Proposition 2.9, [5]), which demonstrates its low-rank property. Motivated by this low-rank property of the (dual) Ky Fan 2- k -norm, which is more general than that of the nuclear norm, and its usage in other applications, in this paper, we propose a Ky Fan 2- k -norm-based non-convex approach for the matrix recovery problem which aims to recover matrices which are not recoverable by the convex relaxation formulation (1). In Section 2, we discuss the proposed models in detail and in Section 3, we develop numerical algorithms to solve those models. Some numerical results will also be presented.

2 Ky Fan 2- k -Norm-Based Models

The Ky Fan 2- k -norm is the ℓ_2 -norm of the vector of k largest singular values with $k \leq \min\{m, n\}$. Thus we have:

$$\|\mathbf{A}\|_{k,2} = \left(\sum_{i=1}^k \sigma_i^2(\mathbf{A}) \right)^{\frac{1}{2}} \leq \|\mathbf{A}\|_F = \left(\sum_{i=1}^{\min\{m,n\}} \sigma_i^2(\mathbf{A}) \right)^{\frac{1}{2}},$$

where $\|\cdot\|$ is the Frobenius norm. Now consider the dual Ky Fan 2- k -norm and use the definition of the dual norm, we obtain the following inequality:

$$\|\mathbf{A}\|_F^2 = \langle \mathbf{A}, \mathbf{A} \rangle \leq \|\mathbf{A}\|_{k,2}^* \cdot \|\mathbf{A}\|_{k,2}.$$

Thus we have:

$$\|\mathbf{A}\|_{k,2} \leq \|\mathbf{A}\|_F \leq \|\mathbf{A}\|_{k,2}^*, \quad k \leq \min\{m, n\}. \quad (4)$$

It is clear that these inequalities become equalities if and only if $\text{rank}(\mathbf{A}) \leq k$. It shows that to find a low-rank matrix \mathbf{X} that satisfies $\mathcal{A}(\mathbf{X}) = \mathbf{b}$ with $\text{rank}(\mathbf{X}) \leq k$, we can solve either the following optimization problem

$$\begin{aligned} \min \quad & \frac{\|\mathbf{X}\|_{k,2}^*}{\|\mathbf{X}\|_F} \\ \text{s.t.} \quad & \mathcal{A}(\mathbf{X}) = \mathbf{b}, \end{aligned} \quad (5)$$

or

$$\begin{aligned} \min \quad & \|\mathbf{X}\|_{k,2}^* - \|\mathbf{X}\|_F \\ \text{s.t.} \quad & \mathcal{A}(\mathbf{X}) = \mathbf{b}. \end{aligned} \quad (6)$$

It is straightforward to see that these non-convex optimization problems can be used to recover low-rank matrices as stated in the following theorem given the norm inequalities in (4).

Theorem 1. *If there exists a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ such that $\text{rank}(\mathbf{X}) \leq k$ and $\mathcal{A}(\mathbf{X}) = \mathbf{b}$, then \mathbf{X} is an optimal solution of (5) and (6).*

Given the result in Theorem 1, the exact recovery of a low-rank matrix using (5) or (6) relies on the uniqueness of the low-rank solution of $\mathcal{A}(\mathbf{X}) = \mathbf{b}$. Recht et al. [11] generalized the restricted isometry property of vectors introduced by Candès and Tao [4] to matrices and use it to provide sufficient conditions on the uniqueness of these solutions.

Definition 1 (Recht et al. [11]). *For every integer k with $1 \leq k \leq \min\{m, n\}$, the k -restricted isometry constant is defined as the smallest number $\delta_k(\mathcal{A})$ such that*

$$(1 - \delta_k(\mathcal{A})) \|\mathbf{X}\|_F \leq \|\mathcal{A}(\mathbf{X})\|_2 \leq (1 + \delta_k(\mathcal{A})) \|\mathbf{X}\|_F \quad (7)$$

holds for all matrices \mathbf{X} of rank at most k .

Using Theorem 3.2 in Recht et al. [11], we can obtain the following exact recovery result for (5) and (6).

Theorem 2. *Suppose that $\delta_{2k} < 1$ and there exists a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ which satisfies $\mathcal{A}(\mathbf{X}) = \mathbf{b}$ and $\text{rank}(\mathbf{X}) \leq k$, then \mathbf{X} is the unique solution to (5) and (6), which implies exact recoverability.*

The condition in Theorem 2 is indeed better than those obtained for the nuclear norm approach (see, e.g., Theorem 3.3 in Recht et al. [11]). The non-convex optimization problems (5) and (6) use norm ratio and difference. When $k = 1$, the norm ratio and difference are computed between the nuclear and Frobenius norm. The idea of using these norm ratios and differences with $k = 1$ has been used to generate non-convex sparse generalizer in the vector case, i.e., $m = 1$. Yin et al. [13] investigated the ratio ℓ_1/ℓ_2 while Yin et al. [14] analyzed

the difference $\ell_1 - \ell_2$ in compressed sensing. Note that even though optimization formulations based on these norm ratios and differences are non-convex, they are still relaxations of ℓ_0 -norm minimization problem unless the sparsity level of the optimal solution is $s = 1$. Our proposed approach is similar to the idea of the truncated difference of the nuclear norm and Frobenius norm discussed in Ma et al [8]. Given a parameter $t \geq 0$, the truncated difference is defined

$$\text{as } \|\mathbf{A}\|_{*,t-F} = \sum_{i=t+1}^{\min\{m,n\}} \sigma_i(\mathbf{A}) - \left(\sum_{i=t+1}^{\min\{m,n\}} \sigma_i^2(\mathbf{A}) \right)^{\frac{1}{2}} \geq 0. \text{ For } t \geq k-1, \text{ the}$$

problem of truncated difference minimization can be used to recover matrices with rank at most k given that $\|\mathbf{X}\|_{*,t-F} = 0$ if $\text{rank}(\mathbf{X}) \leq t+1$. Similar results for exact recovery as in Theorem 2 are provided in Theorem 3.7(a) in Ma et al [8]. Despite the similarity with respect to the recovery results, the problems (5) and (6) are motivated from a different perspective. We are now going to discuss how to solve these problems next.

3 Numerical Algorithm

3.1 Difference of Convex Algorithms

We start with the problem (5). It can be reformulated as

$$\begin{aligned} \max_{\mathbf{Z}, z} \quad & \|\mathbf{Z}\|_F^2 \\ \text{s.t.} \quad & \|\mathbf{Z}\|_{k,2}^* \leq 1, \\ & \mathcal{A}(\mathbf{Z}) - z \cdot \mathbf{b} = \mathbf{0}, \\ & z > 0, \end{aligned} \tag{8}$$

with the change of variables $z = 1/\|\mathbf{X}\|_{k,2}^*$ and $\mathbf{Z} = \mathbf{X}/\|\mathbf{X}\|_{k,2}^*$. The compact formulation

$$\min_{\mathbf{Z}, z} \delta_{\mathcal{Z}}(\mathbf{Z}, z) - \|\mathbf{Z}\|_F^2 / 2, \tag{9}$$

where \mathcal{Z} is the feasible set of the problem (8) and $\delta_{\mathcal{Z}}(\cdot)$ is the indicator function of \mathcal{Z} . The problem (9) is a difference of convex (d.c.) optimization problem (see, e.g. [9]). The difference of convex algorithm DCA proposed in [9] can be applied to the problem (9) as follows.

Step 1. Start with $(\mathbf{Z}^0, z^0) = (\mathbf{X}^0 / \|\mathbf{X}^0\|_{k,2}^*, 1/\|\mathbf{X}^0\|_{k,2}^*)$ for some \mathbf{X}^0 such that $\mathcal{A}(\mathbf{X}^0) = \mathbf{b}$ and set $s = 0$.

Step 2. Update $(\mathbf{Z}^{s+1}, z^{s+1})$ as an optimal solution of the following convex optimization problem

$$\begin{aligned} \max_{\mathbf{Z}, z} \quad & \langle \mathbf{Z}^s, \mathbf{Z} \rangle \\ \text{s.t.} \quad & \|\mathbf{Z}\|_{k,2}^* \leq 1 \\ & \mathcal{A}(\mathbf{Z}) - z \cdot \mathbf{b} = \mathbf{0} \\ & z > 0. \end{aligned} \tag{10}$$

Step 3. Set $s \leftarrow s + 1$ and repeat Step 2.

Let $\mathbf{X}^s = \mathbf{Z}^s/z^s$ and use the general convergence analysis of DCA (see, e.g., Theorem 3.7 in [10]), we can obtain the following convergence results.

Proposition 1. *Given the sequence $\{\mathbf{X}^s\}$ obtained from the DCA algorithm for the problem (9), the following statements are true.*

- (i) *The sequence $\left\{ \frac{\|\mathbf{X}^s\|_{k,2}^*}{\|\mathbf{X}^s\|_F} \right\}$ is non-increasing and convergent.*
- (ii) $\left\| \frac{\mathbf{X}^{s+1}}{\|\mathbf{X}^{s+1}\|_{k,2}^*} - \frac{\mathbf{X}^s}{\|\mathbf{X}^s\|_{k,2}^*} \right\|_F \rightarrow 0$ when $s \rightarrow \infty$.

The convergence results show that the DCA algorithm improves the objective of the ratio minimization problem (5). The DCA algorithm can stop if $(\mathbf{Z}^s, z^s) \in \mathcal{O}(\mathbf{Z}^s)$, where $\mathcal{O}(\mathbf{Z}^s)$ is the set of optimal solution of 10 and (\mathbf{Z}^s, z^s) which satisfied this condition is called a critical point. Note that (local) optimal solutions of (9) can be shown to be critical points. The following proposition shows that an equivalent condition for critical points.

Proposition 2. *$(\mathbf{Z}^s, z^s) \in \mathcal{O}(\mathbf{Z}^s)$ if and only if $\mathbf{Y} = \mathbf{0}$ is an optimal solution of the following optimization problem*

$$\begin{aligned} \min_{\mathbf{Y}} \quad & \|\mathbf{X}^s + \mathbf{Y}\|_{k,2}^* - \frac{\|\mathbf{X}^s\|_{k,2}^*}{\|\mathbf{X}^s\|_F^2} \cdot \langle \mathbf{X}^s, \mathbf{Y} \rangle \\ \text{s.t.} \quad & \mathcal{A}(\mathbf{Y}) = \mathbf{0}. \end{aligned} \quad (11)$$

Proof. Consider $\mathbf{Y} \in \text{Null}(\mathcal{A})$, i.e., $\mathcal{A}(\mathbf{Y}) = \mathbf{0}$, we then have:

$$\left(\frac{\mathbf{X}^s + \mathbf{Y}}{\|\mathbf{X}^s + \mathbf{Y}\|_{k,2}^*}, \frac{1}{\|\mathbf{X}^s + \mathbf{Y}\|_{k,2}^*} \right) \in \mathcal{Z}.$$

Clearly, $\left\langle \frac{\mathbf{X}^s}{\|\mathbf{X}^s\|_{k,2}^*}, \frac{\mathbf{X}^s + \mathbf{Y}}{\|\mathbf{X}^s + \mathbf{Y}\|_{k,2}^*} \right\rangle \leq \left\langle \frac{\mathbf{X}^s}{\|\mathbf{X}^s\|_{k,2}^*}, \frac{\mathbf{X}^s}{\|\mathbf{X}^s\|_{k,2}^*} \right\rangle$ is equivalent to

$$\|\mathbf{X}^s + \mathbf{Y}\|_{k,2}^* - \frac{\|\mathbf{X}^s\|_{k,2}^*}{\|\mathbf{X}^s\|_F^2} \cdot \langle \mathbf{X}^s, \mathbf{Y} \rangle \geq \|\mathbf{X}^s\|_{k,2}^*.$$

When $\mathbf{Y} = \mathbf{0}$, we achieve the equality. We have: $(\mathbf{Z}^s, z^s) \in \mathcal{O}(\mathbf{Z}^s)$ if and only the above inequality holds for all $\mathbf{Y} \in \text{Null}(\mathcal{A})$, which means $f(\mathbf{Y}; \mathbf{X}^s) \geq f(\mathbf{0}; \mathbf{X}^s)$ for all $\mathbf{Y} \in \text{Null}(\mathcal{A})$, where $f(\mathbf{Y}; \mathbf{X}) = \|\mathbf{X} + \mathbf{Y}\|_{k,2}^* - \frac{\|\mathbf{X}\|_{k,2}^*}{\|\mathbf{X}\|_F^2} \cdot \langle \mathbf{X}, \mathbf{Y} \rangle$. Clearly, it is equivalent to the fact that $\mathbf{Y} = \mathbf{0}$ is an optimal solution of (11).

The result of Proposition 2 shows the similarity between the norm ratio minimization problem (5) and the norm different minimization problem (6) with respect to the implementation of the DCA algorithm. It is indeed that the problem (6) is a d.c. optimization problem and the DCA algorithm can be applied as follows.

Step 1. Start with some \mathbf{X}^0 such that $\mathcal{A}(\mathbf{X}^0) = \mathbf{b}$ and set $s = 0$.

Step 2. Update $\mathbf{X}^{s+1} = \mathbf{X}^s + \mathbf{Y}$, where \mathbf{Y} is an optimal solution of the following convex optimization problem

$$\begin{aligned} \min_{\mathbf{Y}} \quad & \|\mathbf{X}^s + \mathbf{Y}\|_{k,2}^* - \frac{1}{\|\mathbf{X}^s\|_F} \cdot \langle \mathbf{X}^s, \mathbf{Y} \rangle \\ \text{s.t.} \quad & \mathcal{A}(\mathbf{Y}) = \mathbf{0}. \end{aligned} \quad (12)$$

Step 3. Set $s \leftarrow s + 1$ and repeat Step 2.

It is clear that \mathbf{X}^s is a critical point for the problem (6) if and only if \mathbf{Y} is an optimal solution of (12). Both problems (10) and (12) can be written in the general form as

$$\begin{aligned} \min_{\mathbf{Y}} \quad & \|\mathbf{X}^s + \mathbf{Y}\|_{k,2}^* - \alpha(\mathbf{X}^s) \cdot \langle \mathbf{X}^s, \mathbf{Y} \rangle \\ \text{s.t.} \quad & \mathcal{A}(\mathbf{Y}) = \mathbf{0}, \end{aligned} \quad (13)$$

where $\alpha(\mathbf{X}) = \frac{\|\mathbf{X}^s\|_{k,2}^*}{\|\mathbf{X}^s\|_F^2}$ for (10) and $\alpha(\mathbf{X}) = \frac{1}{\|\mathbf{X}^s\|_F}$ for (12), respectively. Given that $\mathcal{A}(\mathbf{X}^s) = \mathbf{b}$, this problem can be written as

$$\begin{aligned} \min_{\mathbf{X}} \quad & \|\mathbf{X}\|_{k,2}^* - \alpha(\mathbf{X}^s) \cdot \langle \mathbf{X}^s, \mathbf{X} \rangle \\ \text{s.t.} \quad & \mathcal{A}(\mathbf{X}) = \mathbf{b}. \end{aligned} \quad (14)$$

The following proposition shows that \mathbf{X}^s is a critical point of the problem (14) for many functions $\alpha(\cdot)$ if $\text{rank}(\mathbf{X}^s) \leq k$.

Proposition 3. *If $\text{rank}(\mathbf{X}^s) \leq k$, \mathbf{X}^s is a critical point of the problem (14) for any function $\alpha(\cdot)$ which satisfies*

$$\frac{1}{\|\mathbf{X}\|_F} \leq \alpha(\mathbf{X}) \leq \frac{\|\mathbf{X}^s\|_{k,2}^*}{\|\mathbf{X}^s\|_F^2}. \quad (15)$$

Proof. If $\text{rank}(\mathbf{X}^s) \leq k$, we have: $\alpha(\mathbf{X}^s) = 1/\|\mathbf{X}^s\|_{k,2}$ since $\|\mathbf{X}^s\|_{k,2} = \|\mathbf{X}\|_F = \|\mathbf{X}^s\|_{k,2}^*$. Given that

$$\partial \|\mathbf{A}\|_{k,2}^* = \arg \max_{\mathbf{X} : \|\mathbf{X}\|_{k,2} \leq 1} \langle \mathbf{X}, \mathbf{A} \rangle,$$

we have: $\alpha(\mathbf{X}^s) \cdot \mathbf{X}^s \in \partial \|\mathbf{X}^s\|_{k,2}^*$. Thus for all \mathbf{Y} , the following inequality holds:

$$\|\mathbf{X}^s + \mathbf{Y}\|_{k,2}^* - \|\mathbf{X}^s\|_{k,2}^* \geq \langle \alpha(\mathbf{X}^s) \cdot \mathbf{X}^s, \mathbf{Y} \rangle.$$

It implies $\mathbf{Y} = \mathbf{0}$ is an optimal solution of the problem (13) since the optimality condition is

$$\|\mathbf{X}^s + \mathbf{Y}\|_{k,2}^* - \|\mathbf{X}^s\|_{k,2}^* \geq \langle \alpha(\mathbf{X}^s) \cdot \mathbf{X}^s, \mathbf{Y} \rangle, \quad \forall \mathbf{Y} : \mathcal{A}(\mathbf{Y}) = \mathbf{0}.$$

Thus \mathbf{X}^s is a critical point of the problem (14).

Proposition 3 shows that one can choose different functions $\alpha(\cdot)$ such as $\alpha(\mathbf{X}) = 1/\|\mathbf{X}\|_{k,2}$ for the sub-problem in the general DCA framework to solve the original problem. This generalized sub-problem (14) is a convex optimization problem, which can be formulated as a semidefinite optimization problem given the following calculation of the dual Ky Fan 2- k -norm provided in [5]:

$$\begin{aligned} \|\mathbf{X}\|_{k,2}^* = \min & p + \text{trace}(\mathbf{R}) \\ \text{s.t. } & kp - \text{trace}(\mathbf{P}) = 0, \\ & p\mathbf{I} - \mathbf{P} \succeq 0, \\ & \begin{pmatrix} \mathbf{P} & -\frac{1}{2}\mathbf{X}^T \\ -\frac{1}{2}\mathbf{X} & \mathbf{R} \end{pmatrix} \succeq 0. \end{aligned} \quad (16)$$

In order to implement the DCA algorithm, one also needs to consider how to find the initial solution \mathbf{X}^0 . We can use the nuclear norm minimization problem 1, the convex relaxation of the rank minimization problem, to find \mathbf{X}^0 . A similar approach is to use the following dual Ky Fan 2- k -norm minimization problem to find \mathbf{X}^0 given its low-rank properties:

$$\begin{aligned} \min_{\mathbf{X}} & \|\mathbf{X}\|_{k,2}^* \\ \text{s.t. } & \mathcal{A}(\mathbf{X}) = \mathbf{b}. \end{aligned} \quad (17)$$

This initial problem can be considered as an instance of (14) with $\mathbf{X}^s = \mathbf{0}$ (and $\alpha(\mathbf{0}) = 1$), which is equivalent to starting the iterative algorithm with $\mathbf{X}^0 = \mathbf{0}$ one step ahead. We are now ready to provide some numerical results.

3.2 Numerical Results

Similar to Candès and Recht [3], we construct the following the experiment. We generate \mathbf{M} , an $m \times n$ matrix of rank r , by sampling two $m \times r$ and $n \times r$ factors \mathbf{M}_L and \mathbf{M}_R with i.i.d. Gaussian entries and setting $\mathbf{M} = \mathbf{M}_L \mathbf{M}_R$. The linear map \mathcal{A} is constructed with s independent Gaussian matrices \mathbf{A}_i whose entries follows $\mathcal{N}(0, 1/s)$, i.e.,

$$\mathcal{A}(\mathbf{X}) = \mathbf{b} \Leftrightarrow \langle \mathbf{A}_i, \mathbf{X} \rangle = \langle \mathbf{A}_i, \mathbf{M} \rangle = b_i, \quad i = 1, \dots, s.$$

We generate $K = 50$ matrix \mathbf{M} with $m = 50$, $n = 40$, and $r = 2$. The dimension of these matrices is $d_r = r(m+n-r) = 176$. For each \mathbf{M} , we generate s matrices for the random linear map with s ranging from 180 to 500. We set the maximum number of iterations of the algorithm to be $N_{\max} = 100$. The instances are solved using SDPT3 solver [12] for semi-definite optimization problems in Matlab. The computer used for these numerical experiments is a 64-bit Windows 10 machine with 3.70GHz quad-core CPU, and 32GB RAM. The performance measure is the relative error $\frac{\|\mathbf{X} - \mathbf{M}\|_F}{\|\mathbf{M}\|_F}$ and the threshold $\epsilon = 10^{-6}$ is chosen. We run three different algorithms, **nuclear** used the nuclear optimization formulation (1), **k2-nuclear** used the proposed iterative algorithm with initial solution obtained from (1), and **k2-zero** used the same algorithm with initial solution $\mathbf{X}^0 = \mathbf{0}$.

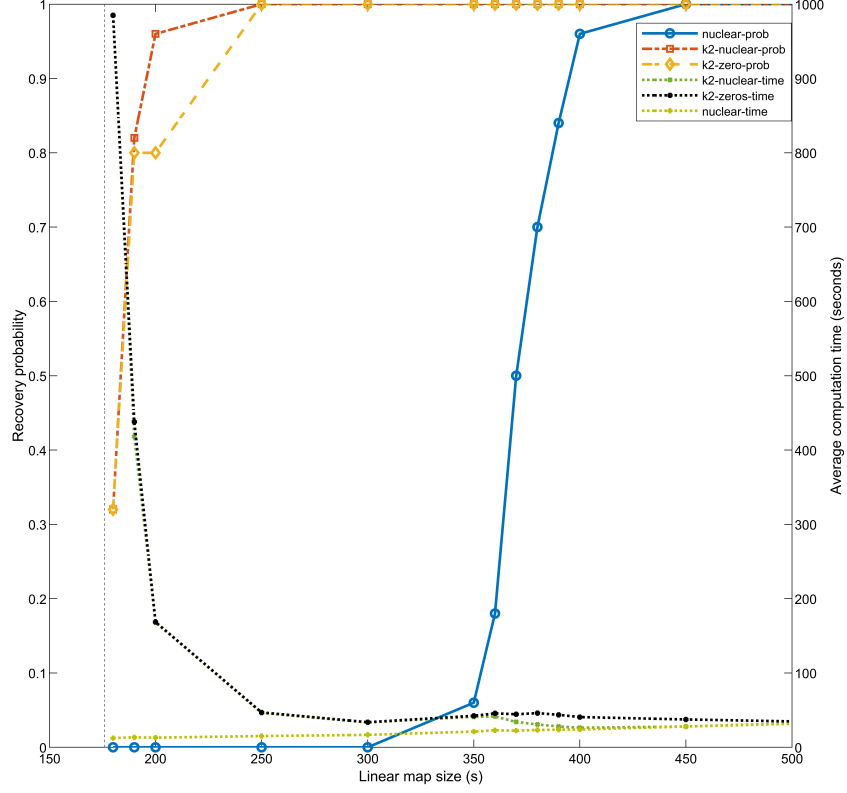


Fig. 1. Recovery probabilities and average computation times of three algorithms

Figure 1 shows recovery probabilities and average computation times (in seconds) for different sizes of the linear map.

The results show that the proposed algorithm can recover exactly the matrix \mathbf{M} with 100% rate when $s \geq 250$ with both initial solutions while the nuclear norm approach cannot recover any matrix at all, i.e., 0% rate, if $s \leq 300$. **k2-nuclear** is slightly better than **k2-zero** in terms of recoverability when s is small while their average computational times are almost the same in all cases. The efficiency of the proposed algorithm when s is small comes with higher average computational times as compared to that of the nuclear norm approach. For example, when $s = 180$, on average, one needs 80 iterations to reach the solution when the proposed algorithm is used instead of 1 with the nuclear norm optimization approach. Note that the average number of iterations is computed for all cases including cases when the matrix \mathbf{M} cannot be recovered. For recov-

erable cases, the average number of iterations is much less. For example, when $s = 180$, the average number of iterations for recoverable case is 40 instead of 80. When the size of the linear map increases, the average number of iterations is decreased significantly. We only need 2 extra iterations when $s = 250$ or 1 extra iteration on average when $s = 300$ to obtain 100% recover rate when the nuclear norm optimization approach still cannot recover any of the matrices (0% rate). These results show that the proposed algorithm achieve significantly better recovery rate with a small number of extra iterations in many cases. We also test the algorithms with higher ranks including $r = 5$ and $r = 10$. Figure 2 shows the results when the size of linear map is $s = \lceil 1.05d_r \rceil$.

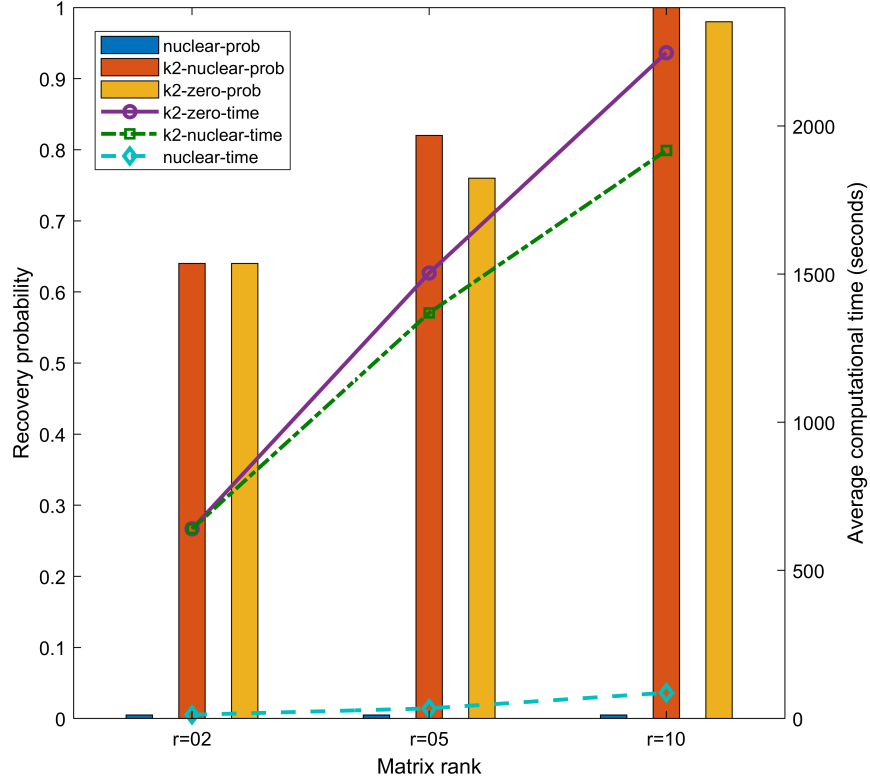


Fig. 2. Recovery probabilities and average computation times for different ranks

These results show that when the size of linear maps is small, the proposed algorithms are significantly better than the nuclear norm optimization approach. With $s = \lceil 1.05d_r \rceil$, the recovery probability increases when r increases and it is

close to 1 when $r = 10$. The computational time increases when r increases given that the size of the sub-problems depends on the size of the linear map. With respect to the number of iterations, it remains low. When $r = 10$, the average numbers of iterations are 22 and 26 for **k2-nuclear** and **k2-zero**, respectively. It shows that **k2-nuclear** is slightly better than **k2-zero** both in terms of recovery probability and computational time.

4 Conclusion

We have proposed non-convex models based on the dual Ky Fan 2- k -norm for low-rank matrix recovery and developed a general DCA framework to solve the models. The computational results are promising. Numerical experiments with larger instances will be conducted with first-order algorithm development for the proposed modes as a future research direction.

References

1. Argyriou, A., Foygel, R., Srebro, N.: Sparse prediction with the k -support norm. In: NIPS. pp. 1466–1474 (2012)
2. Bhatia, R.: Matrix Analysis, Graduate Texts in Mathematics, vol. 169. Springer-Verlag, New York (1997)
3. Candès, E.J., Recht, B.: Exact matrix completion via convex optimization. *Foundations of Computational mathematics* **9**(6), 717–772 (2009)
4. Candès, E.J., Tao, T.: Decoding by linear programming. *IEEE Transactions on Information Theory* **51**(12), 4203–4215 (2005)
5. Doan, X.V., Vavasis, S.: Finding the largest low-rank clusters with Ky Fan 2- k -norm and ℓ_1 -norm. *SIAM Journal on Optimization* **26**(1), 274–312 (2016)
6. Giraud, C.: Low rank multivariate regression. *Electronic Journal of Statistics* **5**, 775–799 (2011)
7. Jacob, L., Bach, F., Vert, J.P.: Clustered multi-task learning: a convex formulation. In: NIPS. vol. 21, pp. 745–752 (2009)
8. Ma, T.H., Lou, Y., Huang, T.Z.: Truncated ℓ_{1-2} models for sparse recovery and rank minimization. *SIAM Journal on Imaging Sciences* **10**(3), 1346–1380 (2017)
9. Pham-Dinh, T., Le-Thi, H.A.: Convex analysis approach to d.c. programming: Theory, algorithms and applications. *Acta Mathematica Vietnamica* **22**(1), 289–355 (1997)
10. Pham-Dinh, T., Le-Thi, H.A.: A d.c. optimization algorithm for solving the trust-region subproblem. *SIAM Journal on Optimization* **8**(2), 476–505 (1998)
11. Recht, B., Fazel, M., Parrilo, P.: Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review* **52**(3), 471–501 (2010)
12. Toh, K.C., Todd, M.J., Tütüncü, R.H.: Sdpt3 - a matlab software package for semidefinite programming, version 1.3. *Optimization Methods and Software* **11**(1-4), 545–581 (1999)
13. Yin, P., Esser, E., Xin, J.: Ratio and difference of ℓ_1 and ℓ_2 norms and sparse representation with coherent dictionaries. *Communications in Information and Systems* **14**(2), 87–109 (2014)
14. Yin, P., Lou, Y., He, Q., Xin, J.: Minimization of $\ell_1 - \ell_2$ for compressed sensing. *SIAM Journal on Scientific Computing* **37**(1), A536–A563 (2015)